

딥러닝 기반의 웨어러블 디바이스에서의 제스처 인식

Gesture recognition with wearable device based on deep learning

□ 변성우, 이석필, 김건년*, 한상현** / 상명대학교, *전자부품연구원, **(주)리딩유아이

요약

본 연구는 비접촉식 센서 기반의 웨어러블 디바이스를 이용한 딥러닝 기반의 제스처 인식에 대한 연구이다. 이를 위하여 Flexible MSG 센서를 기반으로 한 Flexible Epidermal Tactile Sensor를 사용하였으며, Flexible Epidermal Tactile Sensor는 손, 손가락 제스처를 취했을 때 손목, 손가락과 연결되어 있는 근육들의 움직임에 따라 발생하는 피부 표면의 전극을 취득하는 센서이다. 실험을 위하여 7가지 손, 손가락 제스처를 정의하였으며, 손목의 꺾임, 손목의 뒤틀림, 손가락의 오므림과 펴짐, 아무 동작도 취하지 않은 기본 상태에 대한 제스처로 정의하였다. 실험 데이터 수집에는 손목이나 손가락에 부상, 장애등이 없는 일반적인 8명의 참가자가 참가하였으며 각각 한 제스처에 대하여 20번씩 반복하여 1120개의 샘플을 수집하였다. 입력신호에 대한 제스처를 학습하기 위해 본 논문에서는 1차원 Convolutional Neural Network를 제안하였으며, 성능 비교를 위해 신호의 크기를 반영하는 특징벡터인 Integral Absolute Value와 Difference Absolute Mean Value를 입력신호에서 추출하고 Support Vector

Machine을 사용하여 본 논문에서 제안한 1차원 CNN과 성능비교를 하였다. 그 결과 본 논문에서 제안한 1차원 CNN의 분류 정확도가 우수한 성능을 나타냈다.

1. 서론

디지털 방송 서비스가 보급되고, 통신망이 빠르게 발전하면서 IPTV와 같은 양방향 방송 서비스가 가능하게 되었다. 기존의 방송 서버에서만 사용자에게 방송 콘텐츠를 제공하는 방법에서 통신망을 이용하여 사용자의 인터랙션 정보 및 피드백 정보를 방송 서버에 보낼 수 있게 되었다. 또한, 이제는 TV, 오디오 등 많은 주변 공간의 사물들이 네트워크로 연결되면서 이런 장치들과의 인터랙션이 요구되고 있다. 이에 따라, 사용자의 몸 움직임을 이용하여 스마트

※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 기술개발사업의 연구결과로 수행되었음(IITP-2016-R7117-16-0176)

기기들과 직관적으로 쉽게 상호작용하는 제스처 인식 기술에 대한 관심이 커져가고 있다[1].

제스처 인식 기술은 크게 접촉식 방법과 비 접촉식 방법으로 구분할 수 있으며, 접촉식은 센서나 장치를 사용자의 신체에 부착해 이로부터 획득된 데이터를 동작 인식에 활용하는 방식이며, 비 접촉식 방식은 주로 카메라를 이용해 영상으로 사용자의 움직임 정보를 획득하고 이를 활용하여 제스처를 인식한다. 현재 제스처 인식 기술은 대부분 시각 기술에 기반하고 있으며, 하나의 카메라 혹은 여러 대의 카메라들을 통해 획득된 영상으로부터 사용자의 움직임을 추적하고 인식하고 있다[2][3]. 하지만 카메라 영상 기반의 제스처 인식 기술은 주변 환경이나 사용자의 활동반경에 많은 제약을 가지고 있다. 웨어러블 디바이스를 이용한 제스처 인식 기술은 주로 접촉식 방법을 사용하고 있으며, 착용하는 사용자의 움직임을 감지할 수 있는 센서나 장치를 사람의 몸이나 옷에 부착해 데이터를 획득한다. 사용자가 직접 착용해야 하는 불편함은 있지만 센서를 직접 부착하기 때문에 비교적 정확한 동작 정보를 얻을 수 있다. 또한, 비 접촉식과 비교하여, 카메라 시야가 가려진 부분의 제스처는 인식할 수 없는 문제를 해결하여 몸의 자세와 움직임을 어떤 상태에서도 인식할 수 있는 웨어러블 센서를 이용한 제스처 인식 연구가 활발하게 진행되고 있다.

기존의 제스처 인식에 사용된 센서는 스마트 기기의 현재 위치와 이동속도, 흔들림 등을 인지하여 디바이스의 이동 패턴을 산출하는 장치이며, Micro Electro Mechanical Systems(MEMS) 기술을 기반으로 자이로 센서, 가속도 센서 등의 관성 센서가 다양한 동작인식용으로 개발되고 있다. 이밖에도 Electromyography(EMG) 신호, Force Sensing Resistor 센서 등이 이용되고 있으며[4][5], 본 연구

에서는 Flexible Metal Strain Gage(MSG) 센서를 사용하였다.

Hansol Kim 외 6명은 키넥트 스켈레톤 모델을 사용하여 포인팅 제스처를 인식하고 손가락 개수 인식하는 방법을 제안하여 큰 화면 환경에서의 제스처 인터페이스를 개발하였다[6]. 박한훈 외 3명은 키넥스 센서로부터 얻어진 depth 이미지와 color 이미지를 혼합하여 사용하는 손 모양 인식 방법을 제안하였으며 손 영역의 실루엣을 깔끔하게 도출하여 손 모양 인식률을 향상시켰다[7].

한편, 웨어러블 디바이스를 이용한 접촉식 제스처 인식 기술로는 반지 형 제스처 인식 기술이 개발되었는데, Nod사에서 개발한 Nod 링은 검지에 착용한 상태에서 엄지손가락 터치로 클릭하거나, 허공에서 자유롭게 움직여 다양한 제스처 명령을 수행하여 입력장치로서의 유용성을 보였다[8]. Thalmic사에서 개발한 Myo는 암 밴드형 제스처 장치로서, 손의 형태에 따라 변하는 근육의 근전도 신호를 인식하여 손의 포스처 및 제스처를 인식하는 장치를 개발하였다[9]. 하지만, 이러한 장치는 착용 시 마다 근전도 검출 전극의 위치가 변하므로 보정 작업이 필요하고, 사용자마다 근전도 신호가 다르므로 학습과정을 반드시 거쳐야 하는 문제를 가지고 있다.

본 논문에서는 비접촉식 센서 기반의 웨어러블 디바이스를 이용한 딥러닝 기반의 제스처 인식방법을 제안한다. 이를 위하여 Flexible MSG 센서를 기반으로 한 Flexible Epidermal Tactile Sensor를 사용하였다. 실험을 위하여 7가지 손, 손가락 제스처를 정의하였으며, 손목의 꺾임, 손목의 뒤틀림, 손가락의 오므림과 펴짐, 아무 동작도 취하지 않은 기본 상태에 대한 제스처로 정의하였다. 실험 데이터 수집에는 손목이나 손가락에 부상, 장애 등이 없는 일반적인 8명의 참가자가 참가하였으며 각각 한

제스처에 대하여 20번씩 반복하여 1120개의 샘플을 수집하였다. 입력신호에 대한 제스처를 학습하기 위해 본 논문에서는 1차원 Convolutional Neural Network를 제안하였으며, 성능 비교를 위해 신호의 크기를 반영하는 특징벡터인 Integral Absolute Value와 Difference Absolute Mean Value를 입력 신호에서 추출하고 Support Vector Machine을 사용하여 본 논문에서 제안한 1차원 CNN과 성능비교를 하였다.

본 논문의 구성은 다음과 같다. II장에서는 제스처 인식에 사용된 비접촉식 센서에 대한 소개, III장에서는 제안하는 방법, IV장에서는 실험 구성 및 실험 결과를 V장에서 결론으로 끝을 맺는다.

II. Flexible Epidermal Tactile Sensor

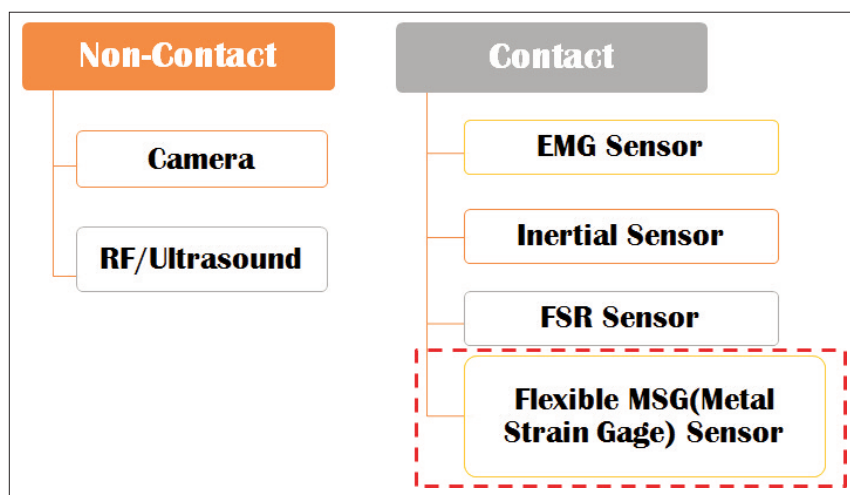
기존의 제스처 인식에 사용된 센서는 MEMS 기

술 기반의 다양한 센서들을 사용하였다.

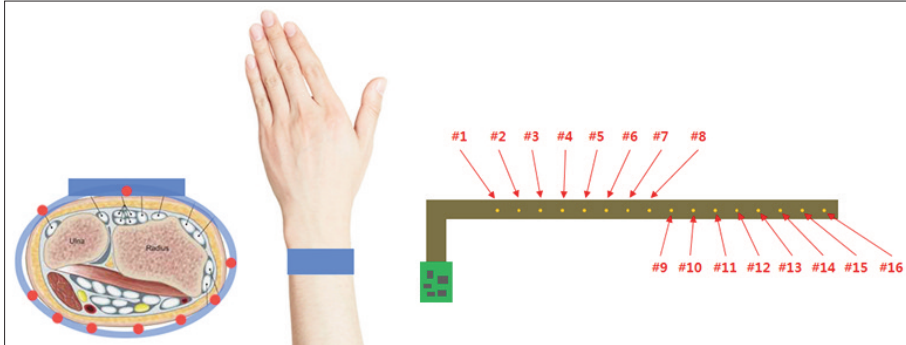
비 접촉식 센서로 카메라, RF/초음파 신호 등을 사용하였으며, 접촉식 센서로는 EMG 센서, 관성 센서, FSR 센서 등을 사용하였다. 본 연구에서는 Flexible MSG 센서를 기반으로 한 Flexible Epidermal Tactile Sensor를 사용하였다.

Flexible MSG 기반의 Flexible Epidermal Tactile Sensor는 손, 손가락 제스처를 취했을 때 손목, 손가락과 연결되어 있는 근육들의 움직임에 따라 발생하는 피부 표면의 전극을 취득하는 센서이다. <그림 2>는 센서의 논리적 모델을 나타낸 그림이다.

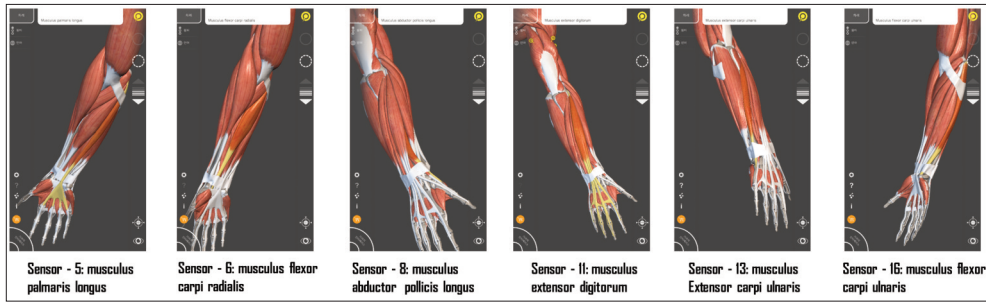
<그림 3>은 센서 번호와 근육을 움직였을 때 반응하는 위치를 매칭한 예시 그림이다. 긴 손바닥근(musculus palmaris longus)이 움직였을 때 발생하는 피부 표면의 전극 신호를 5번 센서에서 취득하고 요골쪽 손목 굽힘근(musculus flexor carpi radialis)이 움직였을 때 발생하는 피부 표면의 전극 신호는 6번 센서에서 취득한다. 이와 같이 각 근육



<그림 1> 제스처 인식을 위한 접촉, 비접촉 센서 기술(Contact and non-contact sensor technologies for gesture recognition)



〈그림 2〉 Flexible Epidermal Tactile Sensor의 논리 모델(Logical model of Flexible Epidermal Tactile Sensor)



〈그림 3〉 근육위치와 센서 번호의 매칭도(Information to map muscle onto the number of sensor)

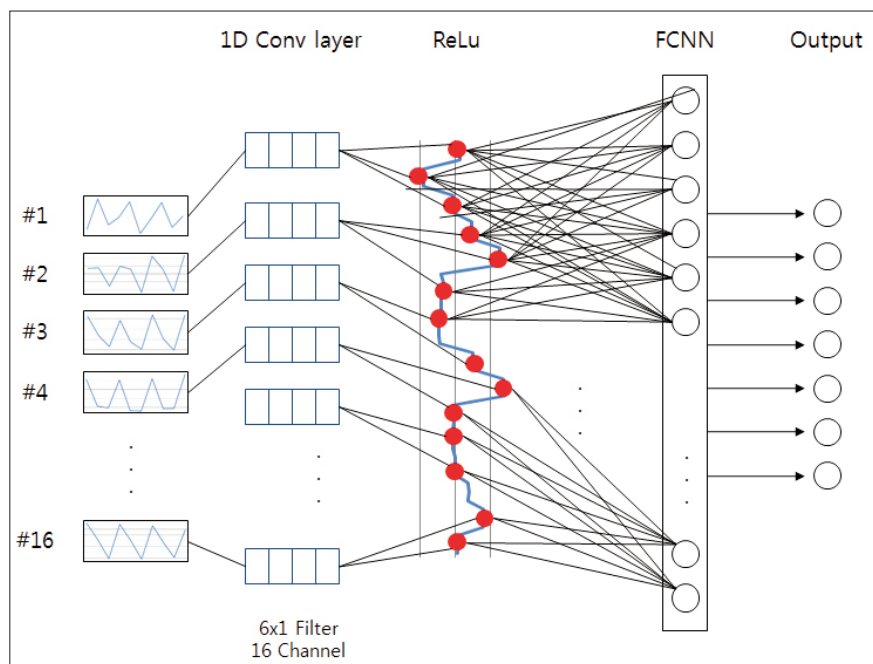
이 움직였을 때 발생하는 피부 표면의 전극을 각 센서에서 취득하게 된다.

III. 1D Convolutional Neural Network

〈그림 4〉는 제안하는 제스처 인식 방법의 구성도이다.

센서의 16개 채널에서 얻어진 표면 전극 신호를 이용하여 손, 손가락 제스처를 인식하기 위해 1차원 Convolutional Neural Network(CNN)을 사용하였다. Neural Network는 훈련을 통해 각각의 가중치 혹은 연결강도를 조정하여 훈련하고 입력의 비선형

함수를 유추할 수 있도록 설계된 머신러닝 알고리즘이다. 일반적으로 CNN에서 입력으로 많이 사용되는 영상이 아닌 1차원의 시계열 신호를 입력으로 받기 때문에 16개 채널의 1차원 Convolution layer를 사용하였고, 웨어러블 기기의 컴퓨팅 환경을 고려하여 구조를 복잡하지 않게 구성하였다. 본 연구에서 사용하는 Neural Network의 구조는 Input layer, 1D Conv layer, LeRu, FCNN, Output layer로 구성되어 있다. Input layer의 입력으로는 16개 채널에서 나온 10개의 시계열 데이터가 들어가게 되고 각 채널의 Filter로 연결되게 된다. Filter의 크기는 6×1 로 실험적으로 정하였으며, Stride 수는 2, 제로 패딩은 0으로 정하였다. Convolution 과정을 거친 신호는 ReLu 활성화 함수를 거쳐서



〈그림 4〉 1차원 Convolutional Neural Network의 구조(Structure of 1D Convolutional Neural Network)

Fully Connected Neural Network(FCNN)로 들어가게 된다. FCNN에도 ReLu 활성화 함수를 사용하였으며, FCNN의 결과로 입력 신호에 대한 제스처 인식 결과를 얻게 된다.

IV. 실험

1. 제스처 정의

본 연구에서는 실험을 위해 7가지 손, 손가락 제스처를 정의하였다.

〈그림 5〉는 제스처 1, 제스처 2에 대한 정의이다. 1번, 2번 제스처는 손목의 꺾임에 대한 제스처로 1번 제스처는 위로 손목 꺾임, 2번 제스처는 아래로 손목 꺾임에 대한 제스처이다.



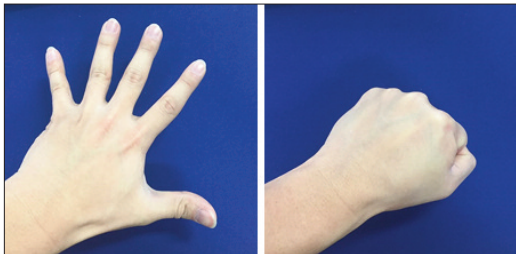
〈그림 5〉 제스처 1(왼쪽) 제스처 2(오른쪽)에 대한 정의(definition for gesture 1(left) and gesture 2(right))

〈그림 6〉은 제스처 3, 제스처 4에 대한 정의이다. 3번, 4번 제스처는 손목의 비틀림에 대한 제스처로 3번 제스처는 위로 손목 비틀림, 4번 제스처는 아래로 손목 비틀림에 대한 제스처이다.

〈그림 7〉은 제스처 5, 제스처 6에 대한 정의이다. 5번, 6번 제스처는 손가락의 오므림과 펴짐에 대한 제스처로 5번 제스처는 손가락 펴짐, 6번 제스처는 손가락 오므림에 대한 제스처이다.



〈그림 6〉 제스처 3(왼쪽) 제스처 4(오른쪽)에 대한 정의(definition for gesture 3(left) and gesture 4(right))



〈그림 7〉 제스처 5(왼쪽) 제스처 6(오른쪽)에 대한 정의(definition for gesture 5(left) and gesture 6(right))



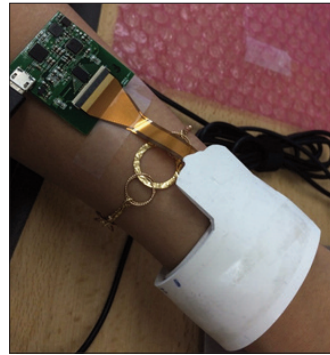
〈그림 8〉 none 상태에 대한 정의(definition for none state)

〈그림 8〉은 제스처 7에 대한 정의이다. 제스처 7은 아무 동작도 취하지 않은 기본 상태에 대한 제스처이다.

2. 실험 데이터

실험데이터를 수집하기 위하여 8명의 참가자가 각각 한 제스처에 대하여 20번씩 반복하였다. 실험 참가자는 손목이나, 손가락 등 손부위에 상처나 장

애가 없는 일반적인 사람이 참가하였으며, 각 제스처 당 160개의 샘플 중 1120개의 샘플을 수집하였다. 그 중 각 제스처 당 100개씩 총 700개의 샘플을 훈련 데이터로 사용하였으며, 각 제스처 당 60개씩 총 420개의 샘플을 테스트 데이터로 사용하였다. 〈그림 9〉는 웨어러블 디바이스 센서를 착용한 모습이다.



〈그림 9〉 웨어러블 디바이스 센서를 착용한 예시(Example to wear sensor for wearable device)

3. 실험 결과

본 논문에서 제안한 1차원 CNN과 성능 비교를 하기 위해 신호의 크기를 반영하는 특징벡터인 Integral Absolute Value(IAV) 특징벡터와 Difference Absolute Mean Value(DAMV) 특징벡터를 사용하였다. IAV 특징벡터는 신호의 절대 적분치를 나타내는 특징으로, 일정시간 동안의 신호의 절대값을 적분한 것이며 식은 다음 식 (1)과 같다.

$$\overline{X} = \sum_{i=1}^N |X(i\Delta t)| \quad (1)$$

DAMV 특징벡터는 각 시계열 신호의 차분 값에

대한 절대 평균치를 나타내는 특징으로 식 (2)와 같다.

$$\bar{X} = \frac{\sum_{i=1}^{N-1} |X(i\Delta t) - X((i+1)\Delta t)|}{N-1} \quad (2)$$

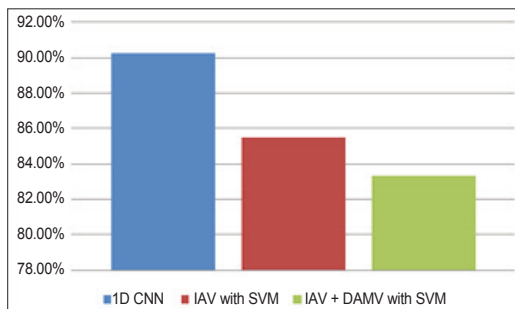
여기에서, X는 측정된 신호 벡터이고, Δt 는 샘플링 시간 간격, N은 샘플의 수이다.

추출한 특징벡터는 가장 일반화를 잘한다고 알려져 있는 머신러닝 알고리즘인 Support Vector Machine(SVM)을 사용하여 분류하고 본 논문에서 제안한 1차원 CNN과 성능 비교하였다.

〈표 1〉은 1차원 CNN의 분류 실험 결과와 IAV 특징벡터를 추출하여 SVM으로 분류한 결과, DAMV 특징벡터를 추출하여 SVM으로 분류한 결과를 비교한 표이다. 1차원 CNN의 분류 정확도는 90.23%로 IAV 특징벡터를 사용했을 때의 정확도 85.47%, IAV와 DAMV 특징벡터를 함께 사용했을 때의 정확도 83.33%보다 우수한 성능을 나타냈다.

〈표 1〉 실험 결과(Experiment results)

	1D CNN	IAV with SVM	IAV + DAMV with SVM
accuracy	90.23%	85.47%	83.33%



〈그림 10〉 실험 결과(Experiment results)

IV. 결론

이제는 TV, 오디오 등 많은 주변 공간의 사물들이 네트워크로 연결되면서 이런 장치들과의 인터랙션이 요구되고 있다. 이에 따라, 사용자의 몸 움직임을 이용하여 스마트 기기들과 직관적으로 쉽게 상호작용하는 제스처 인식 기술에 대한 관심이 커져가고 있다. 웨어러블 디바이스를 이용한 제스처 인식 기술은 주로 접촉식 방법을 사용하고 있으며, 착용하는 사용자의 움직임을 감지할 수 있는 센서나 장치를 사람의 몸이나 옷에 부착해 데이터를 획득한다. 사용자가 직접 착용해야 하는 불편함은 있지만 센서를 직접 부착하기 때문에 비교적 정확한 동작 정보를 얻을 수 있다. 따라서 비 접촉식과 비교하여, 카메라 시야가 가려진 부분의 제스처는 인식할 수 없는 문제를 해결하여 몸의 자세와 움직임을 어떤 상태에서도 인식할 수 있는 웨어러블 센서를 이용한 제스처 인식 연구가 활발하게 진행되고 있다. 본 연구에서는 Flexible MSG 센서를 기반으로 한 Flexible Epidermal Tactile Sensor를 사용하였으며, Flexible Epidermal Tactile Sensor는 손, 손가락 제스처를 취했을 때 손목, 손가락과 연결되어 있는 근육들의 움직임에 따라 발생하는 피부 표면의 전극을 취득하는 센서이다. 실험을 위하여 7가지 손, 손가락 제스처를 정의하였으며, 손목의 꺾임, 손목의 뒤틀림, 손가락의 오므림과 펴짐, 아무 동작도 취하지 않은 기본 상태에 대한 제스처로 정의하였다. 실험 데이터 수집에는 손목이나 손가락에 부상, 장애등이 없는 일반적인 8명의 참가자가 참가하였으며 각각 한 제스처에 대하여 20번씩 반복하여 1120개의 샘플을 수집하였다. 입력신호에 대한 제스처를 학습하기 위해 본 논문에서는 1차원 Convolutional Neural Network를 제안 하였

으며, 성능 비교를 위해 신호의 크기를 반영하는 특징벡터인 Integral Absolute Value와 Difference Absolute Mean Value를 입력신호에서 추출하여 가장 일반화를 잘한다고 알려져 있는 머신 러닝 알고리즘인 Support Vector Machine을 사용하여 분류하고 본 논문에서 제안한 1차원 CNN과 성능비교하였다. 그 결과, 본 논문에서 제안한 1차원 CNN의

분류 정확도는 90.23%로 IAV 특징벡터를 사용했을 때의 정확도 85.47%, IAV와 DAMV 특징벡터를 함께 사용했을 때의 정확도 83.33%보다 우수한 성능을 나타냈다.

향후, 좀 더 정밀한 제스처 인식 기술과 손, 손가락 동작 분석을 위한 연구가 필요하다고 판단된다.

참고 문헌

- [1] Journal of the Institute of Electronics and Information Engineers: Jeong Hyun Tae, Survey on gesture recognition method using wearable device, The Magazine of the IEIE, Vol 42, No. 6, pp. 56-62, 2015, 6
- [2] In proceedings of the 17th International Conference of Pattern Recognition: Y. Yamamoto, I. Yoda, K. sakaue, Arm-pointing gesture interface using surrounded stereo cameras system, ICPR 04, Vol 4, pp. 965-970, 8, 2004
- [3] 5th International Conference on Multimodal Interfaces: Kai, N., Rainer, S, Pointing Gesture Recognition based on 3D-Tracking of Face, Hands and Head Orientation, pp. 140-146, 2003
- [4] The conference of the Institute of Electronics and Information Engineers: Jinhyuk Kim, Shiho Kim, Implementation of Hand Gesture Recognition using wearable Armband EMG sensors, pp. 1121-1122, 6, 2015
- [5] Wearable Computers, 2007 11th IEEE International Symposium on: Georg Ogris, Matthias Kreil, Paul Lukowicz, Using FSR based muscle activity monitoring to recognize manipulative arm gestures, 2007, 10
- [6] Future Information Technology: Hansol Kim, Yoonkyung Kim, Daejune Ko, Jinman Kim, Eui Chul Lee, Pointing Gesture Interface for Large Display Environments Based on the Kinect Skeleton Model, Lecture Notes in Electrical Engineering, Vol 309, pp. 509-514, 2014
- [7] Journal of The Korean Institute of Broadcast and Media Engineers: Hanhoon Park, Junyeong Choi, Jong-II Park, Kwang-Seok Moon, A Study on Hand Region Detection for Kinect-Based Hand Shape Recognition, Vol 18, No. 3, pp. 393-400, 2013, 5
- [8] <http://nod.com>
- [9] <http://www.thalmic.com/myo>

필자소개



변성우

- 2014년 : 상명대학교 디지털미디어학과 이학사
- 2014년 ~ 현재 : 상명대학교 컴퓨터과학과 석박사 연계과정
- 주관심분야 : 멀티미디어 처리, 인공지능

필자 소개



이 석 필

- 1990년 : 연세대학교 전기공학과 공학사
- 1992년 : 연세대학교 전기공학과 공학석사
- 1997년 : 연세대학교 전기공학과 공학박사
- 1997년 ~ 2002년 : 대우전자 영상연구소 선임연구원
- 2002년 ~ 2011년 : KETI 디지털미디어연구센터 센터장
- 2010년 ~ 2011년 : 미국 Georgia Tech. 방문연구원
- 2012년 ~ 현재 : 상명대학교 미디어소프트웨어학과 교수
- 주관심분야 : 멀티미디어 검색, 디지털신호처리, 인공지능



김 건 년

- 2000년 ~ 2003년 : 과기부 국가지정연구실 참여연구원
- 2005년 ~ 2006년 : KETI 나노메카트로닉스 연구센터 센터장
- 2011년 : 고려대학교 전기공학과 공학박사
- 2012년 ~ 2014년 : KETI 차세대융합센서연구센터 센터장
- 2015년 ~ 현재 : KETI 스마트센서 연구센터 수석연구원



한 상 현

- 1996년 : 아주대학교 전자공학과 공학사
- 현재 : 주식회사 리딩유아이 연구소장/CTO